REMARKS

Favorable reconsideration of this application as presently amended and in light of the following discussion is respectfully requested.

Claims 1, 4, 7, 12, 17, 19, 21-31, 34, 35, 39, 43, 50-53, 56, 65, 70, 73-75 and 79-81 are pending in the present application. Claims 8, 20 and 36 have been canceled, Claims 1, 4, 7, 12, 17, 19, 21-23, 25, 27, 31, 39, 43, 53, 56, 65, 73-75 and 79 have been amended and Claim 81 has been added by the present amendment.

In the outstanding Office Action, a copy of the IDS filed on October 5, 1999, was requested; Claims 1, 4, 7, 12, 17, 19-31, 34-36, 39, 43, 50-53, 56, 65, 70 and 73-75 were rejected under 35 U.S.C. § 103(a) as unpatentable over Liddy et al. (hereinafter Liddy) in view of Kishi; and Claims 79 and 80 were allowed.

Applicants thank the Examiner for the indication of allowable subject matter.

Enclosed is a copy of the IDS filed October 5, 1999 as requested in the outstanding Office Action.

Claims 1, 4, 7, 12, 17, 19-31, 34-36, 39, 43, 50-53, 56, 65, 70 and 73-75 stand rejected under 35 U.S.C. § 103(a) as unpatentable over Liddy in view of Kishi. This rejection is respectfully traversed.

Amended Claim 1 is directed to a computer processing apparatus for classifying a document including a database having a database structure providing a classification scheme having a plurality of different subject matter categories. Further, the database contains a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the classification scheme. The database also contains a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group includes a plurality of terms exemplifying the associated category for facilitating

15

disambiguation between different meanings of the same term. The apparatus also includes means for receiving in computer-readable form a document to be classified, processor means for comparing terms appearing in the text document with terms in the database and for determining from the comparison the category for the document, and means for supplying a signal carrying data representing the text document and data associating the text document with the determined category. Independent Claims 31 and 75 include similar features.

On the contrary, Liddy is concerned with a natural language processing system for semantic vector representation which accounts for lexical ambiguity. In Liddy, a lexical database such as the machine-readable tape of the Longman dictionary (LDOCE) is used to allocate subject codes to each word in a text to be processed. The LDOCE is a corpus in which sets of definitions for different senses or meanings of a word are provided and assigned subject codes. In Liddy, a representation of the meaning (context) of unformatted naturally occurring text is generated in the form of subject field codes. The subject field code (SFC) vectors of incoming documents can then be matched to query SFC vectors enabling the documents to be ranked on the basis of similarities.

Thus, in Liddy, after dehypenation, stemming and functional word removal processes, the words in a document to be classified are looked up in the lexical database and the subject code or codes for each word's tagged part of speech is used.

As described at column 7, lines 30-32 of Liddy, a selection of a single subject code is necessary for each word. In other words, where a word could be associated with multiple codes, the codes must be disambiguated. The disambiguation process described in Liddy involves a heuristic order of processes. These processes involve first identifying unique or frequent subject codes. Thus, a computation is made as to whether any subject code in a sentence equals or exceeds a predetermined frequency criterion. If the frequency criterion is exceeded or a subject code is the same as another subject code which was identified as a

16

unique or frequent subject code, then a word is assigned that subject code. However, if

neither the frequency criterion or correspondence to a previously assigned unique or frequent

subject code for the sentence are met, the system in Liddy provides for disambiguation via a

corpus based on subject code correlation. This corpus is a correlation matrix as shown in

Table B of Liddy. The correlation matrix is obtained by correlating pairs of subject field

codes in a corpus of text of the same type as that to be classified by the system. Thus, as can

be seen from Table B, the correlation matrix correlates pairs of subject field codes in a corpus

of text of the same type as that to be classified.

In Liddy, one ambiguous word at a time is resolved by accessing the correlation

matrix via the unique and high frequency subject codes which have been determined for a

sentence containing the word. The system evaluates the correlation coefficients between the

unique frequent subject codes of the sentence and each of the multiple subject codes assigned

to the word being disambiguated to determine which of the multiple subject codes has the

highest correlation with the unique or high frequency subject codes. The system then selects

that subject code as the unambiguous representation of the sense of the word, that is as the

single subject code for the word.

The system described by Liddy is thus very different from the claimed invention.

For example, Liddy does not teach or suggest a database having a database structure

providing a classification scheme having a plurality of different subject matter categories, a

classified vocabulary including a plurality of terms in each of the different subject matter

categories and also a classification data set comprising a plurality of groups of terms with

each group being associated with a specific different one of the subject matter categories and

each group including a plurality of terms exemplifying the associated category for facilitating

disambiguation between different meanings of the same term. Liddy does not use such a

classification data set to facilitate disambiguation between different meanings of the same

17

term. Rather, as described above, Liddy uses a correlation matrix which correlates subject field codes and disambiguates words by accessing the correlation matrix of subject field codes via unique and high frequency subject field codes which have been determined for a sentence containing the ambiguous word and then evaluates the correlation coefficients between the unique frequent subject codes of the sentence and each of the multiple subject codes assigned to the word being disambiguated to determine which of the multiple subject codes has the highest correlation.

Further, Kishi is simply concerned with an automated message processing system configured to automatically manage introduction of the movement of data storage media into a media library.

In addition, amended Claim 12 is directed to a computer processing apparatus for classifying a document including means for accessing a database having a database structure providing a plurality of different subject matter categories. The database contains a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database. The database also contains a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category for disambiguating different meaning of the same term. The apparatus also includes means for receiving in computer-readable form a text document to be classified, processor means for comparing terms appearing in the text document with the collocations to determine the collocation having the most terms in common with the document and for allocating the category of the determined collocation to the document, and means for supplying a signal carrying data representing the text document and data associating the text document with the determined category. Independent Claim 39 is similar to Claim 12, but is a method claim.

Liddy does not teach or suggest a database containing, in addition to a classified

vocabulary, a plurality of collocations with each collocation being associated with a specific

different one of the subject matter categories and each collocation including a plurality of

terms exemplifying the associated category. Rather, as described above, Liddy uses a lexical

database to assign subject field codes to each word in a document and disambiguates between

different meanings of words by using a correlation matrix of subject field codes. Further,

Figure 4 of Liddy simply shows the use of the lexical database to assign subject field codes

while Figure 6 relates to the use of the subject field code correlation matrix to disambiguate

between different meanings of a word. In addition, the description at column 7, line 56-

column 8, line 3 simply suggests that disambiguation may be achieved by selecting, from the

subject codes possible for a word, a subject code which was found to be unique or the most

frequent subject code in similar circumstances by using the correlation matrix.

However, Liddy does not teach or suggest the use of a collocation as in claim 12

where a collocation is a collection of terms that exemplify a category of data and includes

terms which may be used to describe the function, appearance or relationship with other

objects of the classified terms in the associated category or any other terms which may

generally be used in the same context as the classified terms (see, for example, the discussion

of collocations at page 19, line 11 to page 22, line 2 of the specification). The use of these

collocations is conceptually very different from the correlation matrix of Liddy, which

provides information relating the frequency of correlation of certain subject field codes.

Claim 12 does not use the notion of a "predetermined frequency criteria" between subject

field codes, but uses collocations of terms that exemplify subject matter categories to enable

disambiguation. Further, Kishi does not teach or suggest the claimed features.

In addition, amended Claim 65 is directed to an apparatus for classifying electronic

documents including storage means for storing a classification scheme having a plurality of

collocations in which each collocation is associated with a respective different subject matter area and contains a set of terms which exemplify that subject matter area and for facilitating disambiguation between different meanings of the same term. The apparatus also includes means for comparing terms used in a document to be classified with the terms in the collocations, means for allocating the document being classified to the one of the collocations which the comparing means identifies as having the most number of terms in common with the document being classified, means for associating with the document being classified a code representing the subject matter area of the allocated collocation, and means for storing the document together with the associated code.

As described above, Liddy does not teach or suggest the collocations let alone means for comparing terms using a document to be classified with terms in the collocations or means for allocating the document being classified to the one of the collocations which the comparing means identifies as having the most number of terms in common with the document being classified nor means for associating with the document being classified a code representing the subject matter area of the allocated collocations. Rather, in Liddy, each word in the document is allocated a subject field code using a lexical database and, if disambiguation is required, a correlation matrix relating the frequency of different subject field codes is used as set out above. Further, Kishi also does not teach or suggest the features in Claim 65.

Accordingly, in light of the above discussion, it is respectfully submitted independent Claims 1, 12, 31, 39, 65 and 75 patentably define over the combination of Liddy and Kishi.

In addition, new Claim 81 has been added to set forth the invention in a varying scope, and Applicants submit the new claim is supported by the originally filed specification.

Liddy does not teach or suggest a database having a database structure providing a classification scheme, a classified vocabulary and a classification data set comprising a

20

plurality of groups of terms with those groups of terms being provided to facilitate disambiguation between different meanings of the same term let alone of a processor configured to use the groups of terms in the classification data set to disambiguate different meanings of terms in the document and to determine a category for the text document using the database. Rather, in <u>Liddy</u>, disambiguation is effected on the basis of the frequency of occurrence of the subject field codes allocated to words. Thus, if a word is allocated more than one subject field code, a subject field code is selected on the basis of the correlation matrix which correlates frequencies of occurrence of different subject field codes.

Accordingly, it is respectfully submitted independent Claim 81 is also allowable.

Consequently, in light of the above discussion and in view of the present amendment, the present application is believed to be in condition for allowance and an early and favorable action to that effect is respectfully requested.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Customer Number

22850

Tel: (703) 413-3000
Fax: (703) 413 -2220
(OSMMN 08/03)
GJM:DAB/hu
T:\ATTY\DAB\105130001-AM.DOC

Gregory J. Maier
Attorney of Record
Registration No. 25,599
David A. Bilodeau
Registration No. 42,325

22